Sensitivity to Perceived Facial Trustworthiness Is Increased by Activating Self-Protection Motives

Steven G. Young1, Michael L. Slepian2, and Donald F. Sacco3

Abstract
Self-protection motives have been documented to influence a range of intergroup processes, including biased categorization of racially ambiguous targets as out-group members and a heightened ability to discriminate in-group from out-group members. In this work, the influence of self-protective states is extended to interpersonal processes. Specifically, in two experiments we demonstrate that activating self-protection motives (relative to a control experience) leads to more accurate detection of facial cues associated with trustworthiness. In Experiment 1, participants with salient self-protection concerns were better able to distinguish between faces pre-rated as appearing high and low in trustworthiness. In Experiment 2, we used dynamic cues associated with trustworthiness and found that participants with active self-protection goals more accurately distinguished genuine from false smiles. These results are among the first to document the influence of self-protection motives on interpersonal judgments, thereby expanding the scope and focus of fundamental motives research.

Keywords
motivation, self-protection, trustworthiness, social perception

Human cognition and behavior are often regulated and directed by basic, fundamental motives (Kenrick, Neuberg, Griskevicius, Becker, & Schaller, 2010). An integral component of this motivational system involves self-protection, which entails the detection and avoidance of potential sources of danger in the environment, including threats posed by other people (Neuberg, Kenrick, & Schaller, 2011). When activated by environmental cues related to threat, this self-protective motive exerts a top-down influence on a number of perceptual and social cognitive processes. For example, in intergroup contexts, salient self-protection motives are associated with an exacerbated tendency to categorize racially ambiguous targets as out-group members (Maner, Miller, Moss, Leo, & Plant, 2012; Miller, Maner, & Becker, 2010), perceive anger on neutral faces of out-group males (Maner et al., 2005), and engage in more danger-related stereotyping of racial out-groups (Schaller, Park, & Mueller, 2003). Such responses may represent adaptive strategies for addressing safety concerns by prompting the avoidance of targets heuristically considered threatening (e.g., out-group males; Navarrete, McDonald, Molina, & Sidanius, 2010).

In related research, Becker and colleagues (2011) report that individuals for whom self-protection goals are made salient (relative to those in revenge-motive and control conditions) demonstrate greater accuracy in discriminating between faces of friends versus enemies (as indicated by group-identifying insignias briefly presented prior to the faces), results interpreted as evidence of a self-protection vigilance mechanism that tunes perceptual and cognitive processes to threat-relevant information in the environment. This vigilance results in perceptual and attentional enhancements that increase accuracy when distinguishing between friend and foe in intergroup contexts, a critical task for individuals seeking to both avoid others who might pose a threat and approach potential sources of aid and safety.

Self-Protection and Trustworthiness Detection
Although most research has explored how a motive to secure self-protection affects social categorization and intergroup processes, it is likely that self-protective states influence social perception even when intergroup distinctions are not salient. Although a person’s group membership is a useful heuristic to assess their threat potential, other information is often available that may help perceivers concerned with self-protection distinguish between beneficent and threatening conspecifics.
For example, a mere glance at the face of a novel individual affords the opportunity to process structural features (e.g., jaw width, brow prominence, and eye size) that frequently covary with perceived personality traits and characteristics (e.g., Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013). Similarly, dynamic expressions of emotion can suggest internal states and probable future behaviors, including whether a person is likely to act in helpful or harmful ways (e.g., Adams, Ambady, Macrae, & Kleck, 2006; Parkinson, 2005). Therefore, beyond providing information about group membership, the face of a prospective interaction partner offers potentially diagnostic information to a wary perceiver concerned with self-protection.

One such cue to a person’s potential behavioral tendencies that can be signaled both via face structure and emotional displays is trustworthiness. For example, perceptions of trustworthiness based on face structure are formed rapidly and exhibit high levels of inter-rater agreement (Willis & Todorov, 2006). Although there is mixed evidence regarding whether facial appearance predicts actual trustworthiness in interpersonal contexts (Stirrat & Perrett, 2010; cf. Rule, Krendl, Ivcevic, & Ambady, 2013; Todorov, Olivola, Dotsch, & Mendes-Siedlecki, 2015), a person motivated to avoid physical harm is nevertheless likely to seek out individuals who at least appear trustworthy (and are therefore potentially reliable sources of protection). For example, the expression of a smile can signal prosocial intent and positivity when the smile is genuine (i.e., a Duchenne smile), whereas a posed or “social” smile indicates that a person is potentially obscuring their true emotions and intentions (e.g., Ekman, 2003). In fact, faces displaying false smile are perceived as less cooperative and trustworthy than those expressing genuine smiles (Krumhuber et al., 2007) and false smiles are often displayed to hide deception (Biland, Py, Allione, Demarchi, & Abric, 2008).

For a person seeking relief from danger, accurately discriminating between potentially trustworthy and untrustworthy individuals based on these facial signals would be motivationally relevant insofar as it facilitates interacting with others who will offer genuine assistance and not exploit a threatened person’s vulnerability. Of course, there will never be a perfect correspondence between facial cues and actual trustworthiness. Yet, even if the facial cues relied upon to judge trustworthiness are imperfect markers of a persons’ actual disposition and intentions, directing attention to heuristic cues that allow for rapid social decisions still likely proves beneficial on average, much like rapidly discriminating between in-group and out-group members provides coarse but possibly valuable information when seeking protection (e.g., Ekman, 2003).

To begin the experiment, participants were randomly assigned to view one of the two videos on a between-participants basis: a self-protection video or control video. To activate self-protective motives, participants watched an approximately 3-min clip from Silence of the Lambs, while control participants viewed a 3-min clip from Koyaanisqatsi (see Maner et al., 2005 for similar procedure). These clips are well known to induce to self-protection and neutral processes that enhance accurate discrimination of who appears trustworthy or untrustworthy (e.g., based on facial structure) or whose smile suggests genuine beneficence or instead possible dishonesty and obfuscation.

To test this, this research examines how self-protection influences sensitivity to both static and dynamic facial cues that indicate whether a person appears trustworthy or untrustworthy, a distinction of notable importance when determining whether a person is a potential threat or source of protection. Specifically, we examine whether activating self-protective states (compared to control and other negative states) results in perceivers being more sensitive to facial cues associated with the appearance of trustworthiness. In Experiment 1, we manipulate apparent trustworthiness by presenting participants with static images of neutrally expressive target faces that have been found in prior research to appear high and low on the dimension of trustworthiness, based exclusively on the structure of their facial features (Slepian, Young, Rule, Weisbuch, & Ambady, 2012). In Experiment 2, we turn to a dynamic cue to trustworthiness; whether a person displays a genuine or posed smile, which indicates whether they are a safe or potentially disingenuous person to interact with (e.g., Brown & Moore, 2002; Schug, Matsumoto, Horita, Yamagishi, & Bonnet, 2010). We predict that participants with active self-protection goals will more accurately discriminate between individuals who appear trustworthy or untrustworthy, as suggested by both static and dynamic facial information.

**Experiment 1**

**Method**

**Participants**

Fifty-two participants (30 female, $M_{age} = 22.3$ years) completed the experiment in exchange for partial course credit. Two participants were excluded from analyses due to computer malfunctions.

**Procedure**

Upon arrival in the laboratory, participants were greeted by an experimenter, seated in individual cubicles, and informed that they were going to participate in an experiment on social perception. After obtaining informed consent, participants were told that all subsequent instructions would be presented via computer.

To begin the experiment, participants were randomly assigned to view one of the two videos on a between-participants basis: a self-protection video or control video. To activate self-protective motives, participants watched an approximately 3-min clip from Silence of the Lambs, while control participants viewed a neutral clip of equal length depicting time-lapse city-scape images from the film Koyaanisqatsi (see Maner et al., 2005 for similar procedure). These clips are well known to induce to self-protection and neutral
states (Hewig et al., 2005) and have been used previously in similar research (e.g., Becker et al., 2011; Maner et al., 2005). As a cover story, prior to starting the video participants were instructed that, although the video clips were unrelated to subsequent experimental tasks, they would nevertheless be asked questions about the videos at the end of the study and should therefore pay attention.

Next, participants completed a face discrimination task. Here, they viewed 34 male faces that were pre-rated on trustworthiness and separated into trustworthy and untrustworthy sets (with each set containing 17 exemplar faces) based on ratings from prior research (a subset of these faces was used in Slepian et al., 2012). Preliminary tests revealed that the faces differed on trustworthiness, but not other evaluative domains like attractiveness (see Slepian et al., 2012, for additional information). The faces were presented in a random order, shown in gray scale, and displayed in the center of the computer screen until participants rendered a response. For each face, participants indicated whether they believed the person pictured was trustworthy or not (providing a signal detection measure of discriminability). For the purposes of data analysis, we considered a “yes” response to a face pre-rated as trustworthy as a hit, while responding yes to a face pre-rated as untrustworthy was considered a false alarm. Following the trustworthiness discrimination task, participants were thanked for their participation and were debriefed.

Results and Discussion

We first calculated the signal detection measure d’-prime as a measure of participants’ discrimination (d’, e.g., Green & Swets, 1966). Perfect performance (e.g., no false alarms) was corrected using the Snodgrass and Corwin (1988) procedure. To test our prediction that activating self-protective motives would lead to increased discrimination between ostensibly trustworthy and untrustworthy faces, an independent samples t-test compared d’ across conditions and found participants in the self-protection condition demonstrated higher d’ scores ($M = 1.63, SD = 0.65$) than did control participants ($M = 1.19, SD = 0.77$), $t(48) = 2.21, p = .032, d = .62$. Consistent with predictions, active self-protection motives led participants to become more sensitive to whether faces appeared trustworthy or untrustworthy, based on the pretest ratings of the faces.

Although our a priori predictions were specific to accurate discrimination between faces that appeared trustworthy and untrustworthy, we also conducted supplemental analyses of criterion, which allowed us to test whether a self-protective state, relative to control, biased participants to set a higher threshold for perceiving a face as trustworthy. In other words, were participants more inclined to say a face looked “untrustworthy” in the self-protection condition than the control condition? To assess this, we relied on the signal detection statistic $\beta$. However, no difference between the self-protection ($M = 2.85, SD = 1.99$) and control condition ($M = 2.70, SD = 1.93$) was found, $t(48) = 0.28, p = .78$. Although it is difficult to make claims from null results (i.e., here for criterion), the results suggest that self-protective states in the current context do not bias perceivers to see faces as untrustworthy (as an error management perspective might predict; Haselton & Funder, 2006), but rather enhance sensitivity to structural facial features that distinguish between faces that are perceived as trustworthy and untrustworthy.

When in a self-protective state, an adaptive response for humans is to turn to conspecifics for safety and relief (e.g., Schachter, 1959). Face structure is known to provide a heuristic cue to a target’s personality, including whether they appear worthy of our trust (Willis & Todorov, 2006). Although these cues are imperfect indicators of actual personality (cf. Rule et al., 2013; Stirrat & Perrett, 2010), Experiment 1 nevertheless shows that perceiver sensitivity to structural face information perceived as (un)trustworthy is enhanced following the activation of self-protection motives, which might foster distinguishing between conspecifics willing to offer reliable assistance and protection from those who might exploit one’s vulnerability.

Experiment 2

Experiment 1 explored how active self-protective motives moderate sensitivity to structural face information that conveys perceived trustworthiness. Importantly, accuracy in Experiment 1 was determined by agreement with other participants’ pre-ratings of trustworthiness, making claims of increased accuracy in Experiment 1 speculative. Given that perceived trustworthiness of a target has sometimes predicted subsequent behavior (Stirrat & Perrett, 2010), but does not always track actual behavior (Rule et al., 2013), Study 1’s measure should more precisely be labeled as sensitivity to consensual judgments about target faces. Although consensually perceiving others as trustworthy or untrustworthy is still an important social cognitive process (see Todorov et al., 2015), even more germane to the current work would be truly accurate discrimination. We accordingly explored accurate discrimination in Experiment 2 by using stimuli with objective properties more closely associated with trustworthy behavior, not simply a judgment that is ultimately consensus based.

Additionally, while Experiment 1 examined static facial cues, dynamic face information like emotion expressions can also signal a person’s current intentions and trustworthiness (e.g., Parkinson, 2005). For example, a human smile can indicate either a genuine experience of joy or may instead be forced and insincere. One marker of experiencing genuine happiness is the Duchenne smile, which involves the largely involuntary activation of both the zygomaticus major and orbicularis oculi muscles at smile onset, which raise the cheeks and wrinkles the outside corner of the eyes, respectively. Conversely, a non-Duchenne smile recruits only the zygomaticus major and can be produced voluntarily.

These differences in physiognomy and controllability have important implications for the potential trustworthiness of a smiling target (e.g., Ekman, 2003). Duchenne smiles, for
instance, are associated with cooperative behaviors (Mehu, Grammer, & Dunbar, 2007) and can indicate cooperative intentions (Brown & Moore, 2002; Schug et al., 2010). Non-Duchenne smiles, however, are more variable in meaning (e.g., they can be “polite” smiles in social encounters or instead can be used to obfuscate noncooperative intent and untrustworthiness; e.g., Krumhuber et al., 2007). Even if a false smile is not hiding malevolent intent, a non-Duchenne smile does indicate that a person’s emotional displays are regulated and may not reflect their genuine emotional state. With respect to trustworthiness, a false smile suggests a disingenuous and misleading emotional display (e.g., Biland et al., 2008). Consequently, a person seeking protection should be sensitive to whether another person freely reveals their emotions (especially positive and beneficent emotions) or instead is hiding their true emotions. Following this reasoning, we hypothesized that activating self-protective motives would sharpen perceivers’ sensitivity to the distinctions between actual (i.e., not simply perceived) genuine and posed smiles.

Method

Participants

Sixty-five participants (41 female, \(M_{age} = 19.1\) years) completed the experiment in exchange for partial course credit.

Procedure

Upon arrival in the laboratory, participants were greeted by an experimenter and seated in individual cubicles. After securing informed consent, the procedure was similar to that utilized in Study 1, except for the following notable differences. In addition to the self-protection and control videos used in Experiment 1, we included a third condition where participants watched a brief clip taken from The Champ, which is well known to induce sadness (e.g., Hewig et al., 2005). This was done to provide a control condition that was negatively valenced (like the self-protection condition) but that was not expected to influence sensitivity to trustworthiness cues. Participants were randomly assigned to one of the three video conditions on a between-participants basis.

Participants were again told that the videos was unrelated to the face perception task that followed, but that they should pay attention to the videos because they would have to answer questions about them at the end of the study (however, participants were not asked any questions about the videos at the conclusion of the study). After viewing the video clip for the condition they were assigned to, all participants completed the smile discrimination procedure, which served as our primary dependent measure. In this procedure, participants viewed 20 videos of male and female targets, 10 of whom displayed genuine and spontaneous Duchenne smiles while the remaining half displayed non-Duchenne smiles. Stimuli were downloaded from the British Broadcasting Corporation science website (http://www.bbc.co.uk/science/humanbody/mind/surveys/smiles) and have been used in past research (e.g., Bernstein, Young, Brown, Sacco, & Claypool, 2008). Each of the videos (depicting 13 men and 7 women targets) was shown in color, lasted approximately 6–7 s in length, and began by showing a target displaying a neutral expression before smiling and then returning to a neutral expression. Following each video, participants were asked to indicate whether the smile was “real” or “fake.”

We again calculated \(d’\) scores based on these responses, with hits considered correctly calling a Duchenne smile real and false alarms considered incorrectly calling a non-Duchenne smile real. Upon completion of the smile discrimination task, participants were thanked for their participation and debriefed.

Results and Discussion

To examine whether self-protection motives, relative to both a neutral and negative control condition, increased sensitivity to real and posed smiles, we submitted participants’ \(d’\) scores to a one-way analysis of variance, which revealed a significant effect of condition on smile discrimination accuracy, \(F(2, 63) = 3.67, p = .03, \eta^2 = .10\). In order to examine this effect, we next conducted LSD post hoc tests, which found a significant difference when comparing accuracy in the self-protection condition (\(M = 1.31, SD = .73\)) and control condition (\(M = .72, SD = .64\)), \(p = .035\), and a marginal difference between the self-protection and sadness (\(M = .95, SD = .70\)) conditions, \(p = .08\). The control and sadness conditions did not differ, \(p = .27\). As a further test of our hypothesis, a 2, −1, −1, planned contrast comparing smile discrimination in the self-protection condition to the control and sadness conditions found a significant difference, \(t(63) = 2.54, p = .01\). We also conducted supplemental analyses of criterion. Consistent with Experiment 1, we again found no effect of condition on \(b\), as the self-protection (\(M = 1.19, SD = .43\)), control (\(M = 1.23, SD = .64\)), and sadness (\(M = 1.19, SD = .42\)) conditions did not differ, \(F(2, 63) = .053, p = .94\), suggesting that the impact of self-protection motives on real/fake smile discrimination was specific to accuracy and not bias.

In summary, Experiment 2 finds evidence that activating self-protection motives increases sensitivity to whether others are displaying genuine or posed smiles. Additionally, Experiment 2 shows that the influence of self-protective states is not due to a negative mood per se, as inducing sadness did not result in increased accuracy in discriminating between Duchenne and non-Duchenne smiles. Moreover, unlike Experiment 1, this study employed a more objective manipulation (i.e., whether a person honestly displayed emotions or not) that did not rely on unverified consensus judgments of trustworthiness. Finally, rather than using static cues as in Experiment 1, here we manipulated apparent trustworthiness with a dynamic cue, that is, whether a person displayed a smile associated with genuine enjoyment and cooperation (e.g., Mehu et al., 2007) or instead displayed a posed smile, which does not necessarily convey benevolence and suggests an effort to mask one’s genuine emotional state (Ekman, 2003).
General Discussion

Protecting the self from threats, including those posed by other people, is a fundamental social motive (e.g., Kenrick et al., 2010; Miller et al., 2010) and activating self-protection motives strongly influences social cognition. To date, much of the research examining the influence of self-protection motives on social perception and categorization has focused on intergroup relations, documenting the impact of self-protection goals on the perception of out-group emotions (Maner et al., 2005) and perceptual discrimination between members of in-groups and out-groups (e.g., Becker et al., 2011; Maner et al., 2012; Miller et al., 2010). In this work, we show that the social cognitive and perceptual consequences of self-protection motives extend to interpersonal processes. Specifically, we find that self-protective states increase perceivers’ sensitivity to facial cues that suggest trustworthiness (or untrustworthiness). This is a functional response, as threatened individuals would be wise to affiliate with those whose physical appearance and/or facial expressions suggest that they are likely to offer assistance and not exploit their vulnerability.

Notably, this research suggests that the influence of self-protection motivational states is broad, affecting perceiver sensitivity to both static and dynamic facial cues indicative of both consensually based (Experiment 1) and real-word criterion based (Experiment 2) trustworthiness. Finding that perceivers in self-protective states are tuned to aspects of a targets’ face structure and expression that suggest trustworthiness represents a novel contribution to several literatures. First, this is a meaningful addition to a literature that has more commonly emphasized the effect of self-protection on intergroup processes and implies that the self-protective vigilance mechanism proposed by Becker et al. (2011) extends to interpersonal domains, including detecting structural and dynamic facial cues to a targets’ possible traits and underlying motives.

Additionally, the present findings extend the growing literature on motivated person perception. For example, participants whose affiliation motivates is frustrated by experiences and/or facial expressions suggest that they are likely to offer assistance and not exploit their vulnerability. Thus, future investigations demonstrating that self-protection motives increase accurate detection of facial features that more clearly map onto actual traits and behaviors would be valuable. Alternatively, future work could explore how social motives influence the perception of dynamic nonverbal behaviors known to covary with social intentions (paralleling the approach taken in Experiment 2).

Finding a close link between a fundamental motive (the goal to protect oneself) and a fundamental social decision (whether to judge someone as trustworthy or untrustworthy) also proffers speculation about the origins of the ability to rapidly and accurately judge others’ trustworthiness. For example, perhaps the close attunement toward others’ potential trustworthiness in facial structure and expression evolved out of a need to protect oneself upon quick glances of conspecifics. Future work could also examine by which processes these motivational shifts enable more accurate judgments. For instance, perhaps with an enhanced self-protection motive, perceivers saccade more often and more quickly to others’ oribicularis oculi muscle activation (around the eyes) when viewing another person smile in order to determine whether the smile is genuine.

In conclusion, this results present novel evidence that active self-protection motives improve perceivers’ ability to
discriminate between potential allies and foes, in this instance based on both fixed (Experiment 1) and dynamic (Experiment 2) facial cues associated with perceived trustworthiness. We believe that the present studies offer an initial and important demonstration of how fundamental motives modulate a core component of person perception and consider investigating the consequences of self-protective states on person and face perception a potentially fertile area for future research.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References


**Author Biographies**

**Steven G. Young** is an Assistant Professor of Psychology at Baruch College, City University of New York.

**Micheal L. Slepian** is a Postdoctoral Research Scholar at Columbia University. His research examines how people judge others’ trustworthiness, detect deception, and cope with secrecy.

**Donald F. Sacco** is an Assistant Professor of Social Psychology at The University of Southern Mississippi. He received his PhD in Social Psychology from Miami University. His research focuses on face perception, the causes and consequences of social inclusion and exclusion, and the role of motivation in cognition and perception.